--/////////////////////////////////////////////////////////////////////////////////////////////////////////////////////

--checking to see what data looks like when [total\_cases] is null

SELECT \*

FROM [cases & deaths]

WHERE [total\_cases] IS NULL;

--///////////////////////////////////////////////////////////////////

-- Looking through the data, it seems that when [total\_cases] is NULL

-- all other columns are null (except iso\_code & date)

-- this makes sense b/c all other data in the table follows from cases

-- i.e. no deaths, icu or hospital admissions, etc without some number of cases

-- Removing rows from dbo.cases & deaths w/ no useful data

-- where [total cases] is NULL

-- it follows that all other columns are also NULL

BEGIN TRAN;

    BEGIN TRY

        DELETE

        FROM [cases & deaths]

        WHERE [total\_cases] IS NULL

    COMMIT TRAN;

END TRY

BEGIN CATCH

    ROLLBACK TRAN;

END CATCH

--/////////////////////////////////////////////////////////////////////////////////////////////////////////////////

-- Looking at data when [total\_deaths] and [new\_deaths] are NULL

SELECT [iso\_code], [date], [total\_deaths], [new\_deaths]

FROM [cases & deaths]

WHERE [total\_deaths] IS NULL;

SELECT [iso\_code], [date], [total\_deaths], [new\_deaths]

FROM [cases & deaths]

WHERE [total\_deaths] = 0;

SELECT [iso\_code], [date], [total\_deaths], [new\_deaths]

FROM [cases & deaths]

WHERE [total\_deaths] IS NULL AND

    [new\_deaths] IS NOT NULL;

SELECT [iso\_code], [date], [total\_deaths], [new\_deaths]

FROM [cases & deaths]

WHERE [new\_deaths] IS NULL AND

    [total\_deaths] IS NOT NULL;

SELECT [iso\_code], [date], [total\_deaths], [new\_deaths]

FROM [cases & deaths]

WHERE [iso\_code] = 'ZAF' AND

    [date] LIKE '2022-01-%'

-- looks like some [new\_deaths] NULLS should be > 0

SELECT \*

FROM [cases & deaths]

WHERE [new\_deaths] IS NULL;

SELECT \*

FROM [cases & deaths]

WHERE [new\_deaths] = 0;

--////////////////////////////////////////////////////////////////////

-- [total\_deaths] is only NULL when [new\_deaths] is either NULL or 0

-- [total\_deaths] is never 0

-- it is safe to replace [total\_deaths] NULLs with 0

-- [new\_deaths] is NULL when [total\_deaths] > 0

-- in many of those cases the real value should not be 0

-- it is not safe to replace all [new\_deaths] NULLS with 0

-- Replacing all [total\_deaths] NULLs with 0

BEGIN TRAN;

    BEGIN TRY

        UPDATE [cases & deaths]

        SET [total\_deaths] = 0

        WHERE [total\_deaths] IS NULL

    COMMIT TRAN;

END TRY

BEGIN CATCH

    ROLLBACK TRAN;

END CATCH

--///////////////////////////////////////////////////////////////////////////////////////////////////////////////////

-- Finding death rate per cases per country/region

-- Showing only data from most recent date

SELECT loc.location, cases.total\_cases, cases.total\_deaths,

        CAST(CAST(cases.total\_deaths AS FLOAT)/CAST(cases.total\_cases AS FLOAT)\*100 AS decimal(4,2)) AS 'Death\_Rate\_%\_Per\_Cases'

FROM [cases & deaths] AS cases

JOIN [location & population] AS loc

ON cases.iso\_code = loc.iso\_code

WHERE cases.date = '03-29-2022'

ORDER BY 'Death\_Rate\_%\_Per\_Cases' DESC;

-- finding case rate & death rate per population per country/region

SELECT loc.location, loc.population, loc.population\_density,cases.total\_cases, cases.total\_deaths,

        CAST(CAST(cases.total\_cases AS FLOAT)/loc.population\*100 AS decimal(4,2)) AS "Case\_Rate\_%\_Per\_Pop",

        CAST(CAST(cases.total\_deaths AS FLOAT)/loc.population\*100 AS decimal(4,2)) AS "Death\_Rate\_%\_Per\_Pop"

FROM [cases & deaths] AS cases

JOIN [location & population] AS loc

ON cases.iso\_code = loc.iso\_code

WHERE cases.date = '03-29-2022' AND

    loc.population IS NOT NULL

ORDER BY "Case\_Rate\_%\_Per\_Pop" DESC, "Death\_Rate\_%\_Per\_Pop" DESC;

--///////////////////////////////////////////////////////////////////////////////////////////////////////////

SELECT \*

FROM [cases & deaths]